

THE CAUCHY-DAVENPORT THEOREM FOR SEMIGROUPS

SALVATORE TRINGALI

ABSTRACT. We generalize the Davenport transform and use it to prove that, for a (possibly non-commutative) cancellative semigroup $\mathbb{A} = (A, +)$ and non-empty finite subsets X, Y of \mathbb{A} such that the smallest subsemigroup generated by Y is commutative, one has $|X + Y| \geq \Omega(X, Y)$, where

$$\Omega(X, Y) := \min \left(|X| + |Y| - 1, \sup_{y_0 \in Y \times} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0) \right).$$

The result extends the Cauchy-Davenport theorem to the broader setting of semigroups. Also, it strengthens a previous generalization by G. Károlyi relating to sum-sets in a commutative group, for which $\Omega(X, Y)$ in the above inequality is replaced with $\min(p(\mathbb{A}), |X| + |Y| - 1)$, where $p(\mathbb{A})$ is the order of the smallest non-trivial subgroup. While Károlyi's bound is never better than the one provided by $\Omega(X, Y)$, we show that it is in fact much worse in significant cases. Moreover, we prove that our result implies, as a rather immediate corollary, an extension of I. Chowla's generalization of the Cauchy-Davenport theorem to arbitrary cyclic groups.

1. INTRODUCTION

The present paper is focused on the additive theory of semigroups, in continuation to the work initiated in [29], which the reader is recommended to consult for (possibly non-standard) notation and terminology employed here without explanation. For the sake of generality, some definitions and results will, however, be phrased in the more abstract language of magmas, convinced as I am that theorems and theories are better understood (and possibly generalized) in absence of what may be referred to as “conceptual redundancies”. In this respect, let me remark from the outset, for any practical intents and purposes, that all magmas (and hence semigroups, monoids and groups) considered in the sequel, unless differently specified, are written in (standard) additive notation but nonetheless are not commutative. Also, the axiom of choice (shortly, AC) is assumed.

With this in mind, let \mathbb{A} be a magma, that is a pair $(A, +)$ consisting of a (possibly empty) set A and a binary operation $+: A \times A \rightarrow A$. Given two subsets X, Y of \mathbb{A} , we take

$$X + Y := \{x + y : x \in X, y \in Y\}, \quad X - Y := \{z \in A : z + y \in X \text{ for some } y \in Y\}.$$

2000 *Mathematics Subject Classification.* 05E15, 11B13 (primary), 20E99, 20M10 (secondary).

Key words and phrases. Additive theory, Cauchy-Davenport theorem, Davenport transform, groups, product-sets, semigroups, sum-sets, transformation proofs.

The author is funded from the European Community's 7th Framework Programme (FP7/2007-2013) under Grant Agreement No. 276487 (project ApProCEM).

Note that $X + Y = X - Y = \emptyset$ if one of X or Y is empty. In particular, we write $X + y$ in place of $X + Y$ and $X - y$ instead of $X - Y$ whenever $Y = \{y\}$. One calls $X + Y$ and $X - Y$, respectively, the *sum-set* and the *difference-set* of the pair (X, Y) , and then defines $-Y + X$ as the difference set of (X, Y) in the dual, $\mathbb{A}^{\text{op}} := (A, +_{\text{op}})$, of \mathbb{A} , where $+_{\text{op}}$ is the operation $A \times A \rightarrow A : (z_1, z_2) \mapsto z_2 + z_1$.

We say that \mathbb{A} is *unital* if there exists a distinguished element $0_A \in \mathbb{A}$ such that $z + 0_A = 0_A + z = z$ for all z ; when this is the case, 0_A is unique and called the *identity* of \mathbb{A} . Then, we let \mathbb{A}^\times be the set of units of \mathbb{A} , with the convention that $\mathbb{A}^\times := \emptyset$ if \mathbb{A} is not unital; if \mathbb{A} is unital with identity 0_A , a unit of \mathbb{A} is an element $z \in \mathbb{A}$ for which there exists $\tilde{z} \in \mathbb{A}$, provably unique and called the *inverse* of z in \mathbb{A} , such that $z + \tilde{z} = \tilde{z} + z = 0_A$. In addition, for $Z \subseteq \mathbb{A}$ we write $\langle Z \rangle_{\mathbb{A}}$ for the smallest submagma of \mathbb{A} containing Z and $C_{\mathbb{A}}(Z)$ for the center of Z in \mathbb{A} . Lastly, given $z \in \mathbb{A}$ we use $\text{ord}_{\mathbb{A}}(z)$ for the *order* of z in \mathbb{A} , that is $\text{ord}_{\mathbb{A}}(z) := |\langle z \rangle_{\mathbb{A}}|$. The subscript ‘ \mathbb{A} ’ is omitted from the notation whenever \mathbb{A} is implied from the context.

Sum-sets in (mostly commutative) groups have been intensively investigated for several years (see [27] for a recent survey), and interesting results have been also obtained in the case of commutative monoids [12]. The present paper aims to be a further contribute to this popular (and very active) research area, in the direction of extending parts of the theory to the more general setting of semigroups (and magmas), motivated by the apparently reasonable expectation that a further level of abstraction might provide huge benefits.

Historically, the first non-trivial achievement in the study of sum-sets is probably the Cauchy-Davenport theorem, originally established by A. L. Cauchy [2] in 1813, and independently rediscovered by H. Davenport [6, 7] more than a century later. Stated in the wording of group theory (rather than in terms of residue classes, as in the early formulations of Cauchy and Davenport), the theorem reads as follows:

Theorem 1 (The Cauchy-Davenport theorem). *Let \mathbb{A} be a group of prime order p and X, Y non-empty subsets of \mathbb{A} . Then, $|X + Y| \geq \min(p, |X| + |Y| - 1)$.*

The result has been the subject of many papers and much speculation, and has received a number of different proofs, favoring various points of view and eventually leading to significant progress on analogous questions, as in the remarkable case of the Alon-Tarsi’s polynomial method (see [1] and references therein).

The Cauchy-Davenport theorem applies especially to the additive group of the integers modulo a prime. An extension to composite moduli has been very recently given by M. R. Murty and J. P. Whang [21, Theorem 5], based on work by T. C. Tao [28], using tools from Fourier analysis on commutative finite groups. More contributions in the same spirit were previously provided by I. Chowla [4], J. G. van der Corput [5], S. S. Pillai [24], and J. M. Pollard [25]; a partial account of these early results can be found in [22, §2.3], along with an entire chapter dedicated to the celebrated Kneser’s theorem [22, Chapter 4], which implies at once, among the other things, both Theorem 1 and the main result in [4].

Generalizations of a somewhat different flavor have been furnished, still in recent years, by several authors. To be more specific, let 0_A be the identity of a unitization,

$\mathbb{A}^{(1)}$, of a magma \mathbb{A} (as essentially defined by J. M. Howie in [16, p. 2], although in the case of semigroups), and denote by $p(\mathbb{A})$ the cardinality of the smallest submagma of $\mathbb{A}^{(1)}$ generated by z as z ranges in $\mathbb{A}^{(1)} \setminus \{0_A\}$, with the convention that $p(\mathbb{A}) := |\mathbb{N}|$ if $\mathbb{A}^{(1)}$ is trivial, i.e. $\mathbb{A}^{(1)} = \{0_A\}$. Then we have the following theorem, due to G. Károlyi (cf. [17, Theorem 13]):

Theorem 2 (Károlyi’s theorem for commutative groups). *Let \mathbb{A} be a commutative group and X, Y non-empty subsets of \mathbb{A} . Then, $|X + Y| \geq \min(p(\mathbb{A}), |X| + |Y| - 1)$.*

Károlyi’s proof can be classified as a “transformation proof”, and is conceptually similar to other transformation proofs so far invented to deal with the Cauchy-Davenport theorem and related problems in the additive theory of groups; see [10] for details.

While Theorem 2 applies to both finite and infinite *commutative* groups, a like result is known to hold for all *finite* (commutative and non-commutative) groups:

Theorem 3 (Károlyi’s theorem for finite groups). *Let \mathbb{A} be a finite group and X, Y non-empty subsets of \mathbb{A} . Then, $|X + Y| \geq \min(p(\mathbb{A}), |X| + |Y| - 1)$.*

This was first proved by Károlyi himself [18], based on the structure theory of group extensions, by reduction to the case of finite solvable groups in the light of the celebrated Feit-Thompson theorem [11].

One consideration that immediately arises is that, given a finite commutative group \mathbb{A} and non-empty subsets X, Y of \mathbb{A} , the bound for the size of $|X + Y|$ provided by the above Theorems 2 and 3 is far too pessimistic in most situations, as is easily seen, for instance, in the limit case where $X = Y = \mathbb{A}$ and $p(\mathbb{A})$ is somewhat small with respect to the order of \mathbb{A} .

The issue is basically that Theorems 2 and 3 involve a structural “global” property of \mathbb{A} , which is essentially “extraneous” to the pair (X, Y) . Thus, if one really wants to improve Károlyi’s results further, a good idea may be to replace $p(\mathbb{A})$ with something “local”, i.e. more tightly linked to X and Y , which has precisely been the fundamental insight at the origin of this work. Before coming to the main achievements of the paper, we must however recall further results from the literature that are significant with respect to the contents of this manuscript.

The first result is due to J. H. B. Kemperman [19], and deals with torsion-free groups. As with the others above, we report the statement here for the sake of exposition:

Theorem 4 (Kemperman’s inequality for torsion-free groups). *Let \mathbb{A} be a group, and let X, Y be non-empty subsets of \mathbb{A} . Suppose that every non-zero element of \mathbb{A} has order $\geq |X| + |Y| - 1$. Then, $|X + Y| \geq |X| + |Y| - 1$.*

The proof of this, which appears in [19] as a corollary of Theorem 5, proceeds by cleverly iterating what is now sometimes referred to as the Kemperman transform (see, e.g., [10, §2]). It is worth to mention that [19] is, in fact, focused on cancellative semigroups (there simply called semigroups), and it is precisely in this abstract framework that Kemperman establishes a series of results, mostly concerned with the number of different representations of an element in a sum-set, eventually leading to a proof of Theorem 4.

Kemperman's inequality boils down to an instance of Theorem 2 in the special case of *commutative* groups. On another hand, Theorem 4 also represents a major generalization of the following folklore result (whose origins are hard to trace): Suppose X and Y are non-empty subsets of \mathbb{Z} (the ordered ring of integers) of sizes k and ℓ , respectively, and let x_1, x_2, \dots, x_k be a numbering of X and y_1, y_2, \dots, y_ℓ a numbering of Y . Without loss of generality, one can assume, as we do, $x_1 < x_2 < \dots < x_k$ and $y_1 < y_2 < \dots < y_\ell$. Then,

$$x_1 + y_1 < x_2 + y_1 < \dots < x_k + y_1 < x_k + y_2 < \dots < x_k + y_\ell,$$

with the result that $|X + Y| \geq k + \ell - 1$. Interestingly, these same considerations go through almost verbatim in the case of linearly orderable magmas (see [29, Proposition 3.2]), apart from minor complications due to the introduction of parenthesizations (see Definition 2):

Proposition 1.1. *Suppose $\mathbb{A} = (A, \cdot)$ is a linearly ordered magma, written multiplicatively. Pick an integer $n \geq 1$, and let X_1, X_2, \dots, X_n be non-empty subsets of \mathbb{A} and $r_i := |X_i|$. Then, for any central parenthesization P of \mathbb{A} of length n , one has that*

$$(1) \quad |(X_1 X_2 \cdots X_n)_P| \geq 1 - n + \sum_{i=1}^n r_i.$$

Furthermore, (1) is sharp whenever \mathbb{A} is a semigroup and X_1, X_2, \dots, X_n are finite, the lower bound being attained by taking, for instance, $X_i = \{x^k : k \in \mathbb{N}^+, k \leq r_i\}$ for each i , where x is a suitable element of \mathbb{A} (not dependent on i).

As for the rest, S. Eliahou and M. Kervaire [9, Theorem 2.1], improving an earlier result of S. Yuzvinsky linked to the Hurwitz problem in topology [32], have established a variant of Theorem 1 for sum-sets in vector spaces over finite fields. Also, a Cauchy-Davenport theorem for acyclic unital semigroups has been given by the late Y. O. Hamidoune and coauthors [3, Theorem 3], while a graph-theoretic analogue has been recently proved by P. Hegarty [15].

In fact, we are not aware of more results that can be properly regarded as “natural” variants or generalizations of the Cauchy-Davenport theorem, and indeed of no previous work concerned with an extension of the theorem to the setting of arbitrary semigroups.

Thus, with this background in mind, we can finally proceed to state the main result of the paper, but first we need the following definition.

Definition 1. Given a magma $\mathbb{A} = (A, +)$ and a subset Z of \mathbb{A} , we let

$$(2) \quad \omega(Z) := \sup_{z_0 \in Z \times} \min_{z \in Z \setminus \{z_0\}} \text{ord}(z - z_0),$$

Then, for $X, Y \subseteq \mathbb{A}$ we define $\Omega_{\mathbb{A}}(X, Y) := 0$ if $X = \emptyset$ or $Y = \emptyset$; $\Omega_{\mathbb{A}}(X, Y) := \max(|X|, |Y|)$ if (at least) one of X or Y is infinite, and $\Omega_{\mathbb{A}}(X, Y) := \min(\omega(Y), |X| + |Y| - 1)$ otherwise. We call $\Omega_{\mathbb{A}}(X, Y)$ the *Cauchy-Davenport constant of (X, Y) relative to \mathbb{A}* , and write it simply as $\Omega(X, Y)$ when \mathbb{A} is understood from the context.

Remark 1. We assume here that the supremum and maximum of the empty set are 0, while its infimum and minimum are ∞ (by which we mean, as usual, something

larger than any given cardinal). Thus, since a supremum over a non-empty finite set of cardinal numbers is actually a maximum, we can obviously replace \sup with \max in (2) whenever $1 \leq |Z^\times| < \infty$. Perhaps more interestingly, something similar applies to \min and \inf , which looks intriguing, or at least suggestive, inasmuch as it relates, as we are going to see, the problem of sizing a sum-set to a sup-inf condition. And we cannot really refrain from wondering whether there is something more to disclose behind the curtain in this respect.

Every pair (X, Y) of subsets of a given magma \mathbb{A} has a well-defined Cauchy-Davenport constant, which is zero (essentially by definition) if $Y^\times = \emptyset$. However, this is not the case, e.g., when \mathbb{A} is a group and Y is non-empty, to the effect that $\Omega(X, Y)$ can be used for a non-trivial lower bound on the size of $X + Y$ as in the next:

Theorem 5. *Let \mathbb{A} be a cancellative semigroup and X, Y subsets of \mathbb{A} such that $\langle Y \rangle_{\mathbb{A}}$ is commutative. Then $|X + Y| \geq \Omega(X, Y)$.*

Theorem 5 represents the most relevant contribution of the paper: Not only it extends Theorem 1 to the more general and abstract setting of semigroups (see Section 5). It also provides a strengthening of Theorem 2, in that, given a cancellative unital semigroup $\mathbb{A} = (A, +)$ and a subset Z of \mathbb{A} with $Z^\times \neq \emptyset$, it is found (see Lemma 5.1) that

$$\sup_{z_0 \in Z^\times} \min_{z \in Z \setminus \{z_0\}} \text{ord}(z - z_0) \geq p(\mathbb{A}),$$

with strict inequality in some significant cases (see Examples 1 and 2). Theorem 5 is proved in Section 4. The argument is a delicate refinement of the transformation proof originally used by Davenport in [6]. This leads us to define what we call a *generalized Davenport transform*, which might perhaps be considered an interesting by-product in its own right.

Indeed, I am not aware of any previous use of the same technique in a non-commutative setting, much less in relation to (cancellative) semigroups. With few exceptions, remarkably including A. G. Vosper's original proof of his own famous theorem on critical pairs [30], the “Davenport transform” seems, in fact, to have been more or less forgotten, in favor of conceptually similar (but substantially different) “technology” such as the Dyson transform [22, p. 42] or the aforementioned Kemperman transform [19].

Remark 2. One thing seems worth mentioning before proceeding: While it is true that every *commutative* cancellative semigroup embeds as a subsemigroup into a group, this is false in the non-commutative case; see [20] for an explicit example. This serves as a fundamental motivation for the present paper, in that it shows that the study of sum-sets on cancellative semigroups cannot be systematically reduced, in the absence of commutativity, to the case of groups, at least not in any obvious way.

Based on Theorem 5, we provide two proofs of Chowla's extension of the Cauchy-Davenport theorem to arbitrary cyclic groups [4], the second of which comes as a part of the following generalization of Chowla's original result:

Theorem 6. *For $m \in \mathbb{N}^+$ take X and Y to be non-empty subsets of the additive group $\mathbb{A} = (\mathbb{Z}/m\mathbb{Z}, +, -, 0_m)$ of the integers modulo m . Let x_1, x_2, \dots, x_k be a numbering of*

X and y_1, y_2, \dots, y_ℓ a numbering of Y . Define

$$\delta_x := \min_{1 \leq i \leq k} \max_{1 \leq j \leq k, j \neq i} \gcd(m, x_i - x_j), \quad \delta_y := \min_{1 \leq i \leq \ell} \max_{1 \leq j \leq \ell, j \neq i} \gcd(m, y_i - y_j)$$

(by identifying a residue class with any of its representatives), and set $\delta := \max(\delta_x, \delta_y)$. Then

$$|X + Y| \geq \min(\delta^{-1}m, |X| + |Y| - 1).$$

In particular, $|X + Y| \geq \min(m, |X| + |Y| - 1)$ if there exists $y_0 \in Y$ (respectively, $x_0 \in X$) such that m is prime with $y - y_0$ for each $y \in Y \setminus \{y_0\}$ (respectively, prime with $x - x_0$ for each $x \in X \setminus \{x_0\}$).

Many natural questions arise. In particular, one can ask if it is possible to generalize Theorem 5 in such a way to get rid of the assumption that $\langle Y \rangle_{\mathbb{A}}$ is commutative. Unfortunately, at present, I have no satisfactory answer in this respect. However, the question looks interesting, since an affirmative response would provide a comprehensive generalization of about all the previous extensions of the Cauchy-Davenport theorem recalled in this introduction, and most remarkably of Theorems 3 and 4.

1.1. Organization. The plan of the paper is as follows. In Section 2 we establish basic identities and estimates concerning sum-sets in semigroups and magmas. In Section 3 we introduce the generalized Davenport transform and prove some of its fundamental properties. In Section 4 we prove our main theorem. In Section 5 we provide some additional results, including proofs of Theorem 2, Theorem 6 and Chowla's theorem for cyclic groups [4]. In Section 6 we use Theorem 5, in combination with Hall's theorem on distinct representatives, to gain some more information on the “internal arrangement” of sum-sets (by adapting a previous idea of Ø. J. Rødseth [26]). Lastly, in Section 7 we include a couple of examples to show that Theorem 5 can be significantly sharper than Theorem 2.

2. PRELIMINARIES

This section collects basic properties of sum-sets that will be used later to introduce the generalized Davenport transform and prove Theorem 5. Some proofs are direct and standard (and thus omitted without further explanation), but we have no explicit references to anything similar in the context of semigroups (and magmas), so we include the statements here for the sake of exposition. But first we recall the notion of a central parenthesization from [29, §2.1].

Definition 2. Let $\mathbb{A} = (A, \star)$ be a magma. Given $n \in \mathbb{N}^+$, we define recursively $\mathcal{P}_1(\mathbb{A}) := \{\text{id}_A\}$ and $\mathcal{P}_{n+1}(\mathbb{A}) := \mathcal{L}_{n+1}(\mathbb{A}) \cup \mathcal{R}_{n+1}(\mathbb{A})$, where id_A is the map $A \rightarrow A : a \rightarrow a$ and

- (i) $\mathcal{L}_{n+1}(\mathbb{A})$ is the set of all functions $\mathbb{A}^{n+1} \rightarrow \mathbb{A}$ sending, for some $f \in \mathcal{P}_n(\mathbb{A})$, a $(n+1)$ -tuple $(a_1, a_2, \dots, a_{n+1})$ to the product $a_1 \star f(a_2, a_3, \dots, a_{n+1})$.
- (ii) $\mathcal{R}_{n+1}(\mathbb{A})$ is the set of all functions $\mathbb{A}^{n+1} \rightarrow \mathbb{A}$ mapping, for some $f \in \mathcal{P}_n(\mathbb{A})$, a $(n+1)$ -tuple $(a_1, a_2, \dots, a_{n+1})$ to the product $f(a_1, a_2, \dots, a_n) \star a_{n+1}$.

For $n \in \mathbb{N}^+$, we refer to an element P of $\mathcal{P}_n(\mathbb{A})$ as a *central parenthetization* of \mathbb{A} of length n , or also a *central n -parenthetization* of \mathbb{A} . Moreover, for $a_1, a_2, \dots, a_n \in \mathbb{A}$, we write $(a_1 \star a_2 \star \dots \star a_n)_P$ in place of $P(a_1, a_2, \dots, a_n)$ and, whenever S_1, S_2, \dots, S_n are subsets of \mathbb{A} , we let

$$(S_1 \star S_2 \star \dots \star S_n)_P := \{(s_1 \star s_2 \star \dots \star s_n)_P : s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n\}$$

(note that this is empty if $S_i = \emptyset$ for some i). When $S_1 = S_2 = \dots = S_n =: S$, we use $(S^n)_P$ in place of $(S_1 \star S_2 \star \dots \star S_n)_P$ if \mathbb{A} is written multiplicatively, and $(nS)_P$ if it is written additively. Moreover, if \mathbb{A} is a semigroup or $n \leq 2$, then we simply write $S_1 \star S_2 \star \dots \star S_n$ in place of $(S_1 \star S_2 \star \dots \star S_n)_P$, since in both cases the result does not really depend on the choice of P . Similar considerations apply to $(a_1 \star a_2 \star \dots \star a_n)_P$, in the most obvious way.

Parenthetization is just a formal way to deal with expressions involving three or more operands in a magma whose operation is not associative. Also, note that we are not considering here all of the possible parenthesizations of such expressions, but only some special class of them (which we refer to as central). Based upon these premises, we can proceed with a couple of lemmas, providing trivial identities and estimates (cf. [28, Lemma 2.1]).

Lemma 2.1. *Let X, Y, X_1, X_2, Y_1, Y_2 be subsets of a magma $\mathbb{A} = (A, +)$ for which $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$. Then the following holds:*

- (i) $X + Y = Y +_{\text{op}} X$, and hence $|X + Y| = |Y +_{\text{op}} X|$.
- (ii) $X_1 + Y \subseteq X_2 + Y$ and $X + Y_1 \subseteq X + Y_2$.
- (iii) $|X_1 + Y| \leq |X_2 + Y|$ and $|X + Y_1| \leq |X + Y_2|$.

In spite of being so trivial, Lemma 2.1 is useful in many situations, for instance to prove that a certain property holds for semigroups that are left or right cancellative by just proving that it holds in one of the two cases, which allows us to cut by half the number and the length of some mathematical statements. In particular, to express that something is true by point (i) of Lemma 2.1, we will simply say that it is true “by duality”.

Lemma 2.2. *For a left cancellative magma $\mathbb{A} = (A, +)$ the following holds:*

- (i) $|X| = |z + X|$ for every $z \in \mathbb{A}$ and $X \subseteq \mathbb{A}$.
- (ii) If X, Y are subsets of \mathbb{A} and $X \neq \emptyset$, then $|Y| \leq |X + Y|$.
- (iii) Given an integer $n \geq 1$, let X_1, X_2, \dots, X_n be non-empty subsets of \mathbb{A} and P a central parenthesization of \mathbb{A} of length n . One then has that $|X_n| \leq |(X_1 + X_2 + \dots + X_n)_P|$.

Proof. We prove point (iii). For set $X := (X_1 + X_2 + \dots + X_n)_P$. The claim is obvious if $n = 1$, while for $n = 2$ it reduces to point (ii) above. Therefore, let $n \geq 3$ and suppose that the statement holds true for every $k = 1, 2, \dots, n - 1$. There exists a central parenthesization Q of \mathbb{A} of length $n - 1$ such that $X = (X_1 + \dots + X_{n-1})_Q + X_n$ or $X = X_1 + (X_2 + \dots + X_n)_Q$. Since the sets X_1, X_2, \dots, X_n are non-empty, the same point (ii) above and the inductive hypothesis then give that $|X_n| \leq |X|$ in the first case,

and $|X_n| \leq |(X_2 + \cdots + X_n)_Q| \leq |X|$ in the second. Thus, the conclusion follows (by induction). \blacksquare

For the subsequent lemma we assume that $0 \cdot \kappa := \kappa \cdot 0 := 0$ for every cardinal number κ .

Lemma 2.3. *Let $\mathbb{A} = (A, +)$ be an arbitrary magma. One has the following:*

- (i) *If X, Y are subsets of \mathbb{A} , then $|X + Y| \leq |X| \cdot |Y|$.*
- (ii) *Given an integer $n \geq 1$, let X_1, X_2, \dots, X_n be subsets of \mathbb{A} and P a central parenthesization of \mathbb{A} of length n . Then $|(X_1 + X_2 + \cdots + X_n)_P| \leq \prod_{k=1}^n |X_k|$.*

Proof. We prove point (ii). For set $X := (X_1 + X_2 + \cdots + X_n)_P$. For $n = 1$ the assertion is obvious, while for $n = 2$ it reduces to point (i) above. As a consequence, assume $n \geq 3$ and let the statement hold true for every $k = 1, 2, \dots, n - 1$. There exists a central parenthesization Q of \mathbb{A} of length $n - 1$ such that $X = (X_1 + \cdots + X_{n-1})_Q + X_n$ or $X = X_1 + (X_2 + \cdots + X_n)_Q$. It follows from point (ii) that $|X| \leq |(X_1 + \cdots + X_{n-1})_Q| \cdot |X_n|$ in the first occurrence and $|X| \leq |X_1| \cdot |(X_2 + \cdots + X_n)_Q|$ in the second. Whatever the case may be, the inductive hypothesis then gives that $|X| \leq \prod_{k=1}^n |X_k|$, which implies the claim (by induction). \blacksquare

Remark 3. No matter if the ambient semigroup is cancellative, nothing similar to Lemmas 2.2 and 2.3 applies, in general, to difference-sets of type $X - Y$, to the extent that $X - Y$ can be empty even if X and Y are infinite. On another hand, it follows, by duality, from point (i) of Lemma 2.2 that, in presence of cancellativity, the cardinality of sum-sets of type $X + Y$ is preserved under translation, which is a point in common with the case of groups, save the fact that one cannot take advantage of this, at least in general, to “normalize” either of X or Y in such a way as to contain some distinguished element of \mathbb{A} .

The next lemma is central to the use of Davenport transforms in our proof of Theorem 5.

Lemma 2.4. *Let \mathbb{A} be a semigroup and X, Y subsets of \mathbb{A} . Then, the following conditions are equivalent to each other:*

- (i) $X + 2Y \subseteq X + Y$.
- (ii) $X + nY \subseteq X + Y$ for all $n \in \mathbb{N}^+$.
- (iii) $X + \langle Y \rangle_{\mathbb{A}} = X + Y$.

Proof. Points (ii) and (iii) are clearly equivalent, as $X + \langle Y \rangle_{\mathbb{A}} = \bigcup_{n=1}^{\infty} (X + nY)$, and (i) is obviously implied by (ii). Thus, we are left to prove that (ii) follows from (i), which is immediate, by a routine induction, using the fact that, if $X + nY \subseteq X + Y$ for some $n \in \mathbb{N}^+$, then $X + (n + 1)Y = (X + nY) + Y \subseteq (X + Y) + Y = X + 2Y \subseteq X + Y$. \blacksquare

On another hand, the following lemma shows that, in reference to Theorem 5, there is no loss of generality in assuming that the ambient semigroup is unital.

Lemma 2.5. *Suppose that $\mathbb{A}_1 = (A_1, +_1)$ and $\mathbb{A}_2 = (A_2, +_2)$ are magmas, and let $\phi : \mathbb{A}_1 \rightarrow \mathbb{A}_2$ be a (magma) monomorphism and X, Y subsets of \mathbb{A}_1 . Then $|X +_1 Y| = |\phi(X) +_2 \phi(Y)|$.*

We close this section with few simple but remarkable properties of units:

Lemma 2.6. *Let $\mathbb{A} = (A, +)$ be a unital semigroup with identity 0_A , X a subset of \mathbb{A} , and z a unit of \mathbb{A} with inverse \tilde{z} . Then the following holds:*

- (i) $X - z = X + \tilde{z}$ and $-z + X = \tilde{z} + X$.
- (ii) If \mathbb{A} is cancellative, then $|-z + X| = |X - z| = |X|$.
- (iii) If $z \in C_{\mathbb{A}}(X)$ and \mathbb{A} is cancellative, then $\tilde{z} \in C_{\mathbb{A}}(X)$, and in addition $\langle X - z \rangle_{\mathbb{A}}$ and $\langle -z + X \rangle_{\mathbb{A}}$ are commutative if $\langle X \rangle_{\mathbb{A}}$ is commutative.

Proof. (i) By duality, it suffices to prove that $X - z = X + \tilde{z}$. But this is trivial from the fact that $w \in X - z$ if and only if there exists $x \in X$ such that $w + z = x$, which in turn is clearly equivalent to saying that $x + \tilde{z} = w + z + \tilde{z} = w$, viz $w \in X + \tilde{z}$.

(ii) It is straightforward by duality, point (i) of Lemma 2.2 and point (i) above.

(iii) Suppose $z \in C_{\mathbb{A}}(X)$ and pick $x \in X$. By cancellativity, one has that $x + \tilde{z} = \tilde{z} + x$ if and only if $x = \tilde{z} + x + z$, and this condition is certainly verified as $\tilde{z} + x + z = \tilde{z} + z + x = x$ (from the hypothesis that z is in the center of \mathbb{A}). It follows that $\tilde{z} \in C_{\mathbb{A}}(X)$. With this in hand, assume that $\langle X \rangle_{\mathbb{A}}$ is commutative and let $w_1, w_2 \in \langle X - z \rangle_{\mathbb{A}}$. By point (i) above, there must exist $x_1, x_2 \in X$ such that $w_i = x_i + \tilde{z}$, which easily implies that

$$w_1 + w_2 = x_1 + \tilde{z} + x_2 + \tilde{z} = x_1 + x_2 + 2\tilde{z} = x_2 + x_1 + 2\tilde{z} = x_2 + \tilde{z} + x_1 + \tilde{z} = w_2 + w_1,$$

using that $\tilde{z} \in C_{\mathbb{A}}(X)$, as we have just proved, and $\langle X \rangle_{\mathbb{A}}$ is commutative. This ultimately shows that $\langle X - z \rangle_{\mathbb{A}}$ is commutative too, which completes the proof by duality. \blacksquare

Remark 4. There is a notational subtleness here that it may be worth to underline before proceeding. Suppose that \mathbb{A} is a unital semiring and $x, y \in \mathbb{A}$. In principle, $x - y$ and $-y + x$ are not elements of \mathbb{A} : In fact, they are difference-sets, and no other meaningful interpretation is (a priori) possible. However, if $y \in \mathbb{A}^\times$ and \tilde{y} is the inverse of y in \mathbb{A} , then $x - y = \{x + \tilde{y}\}$ and $-y + x = \{\tilde{y} + x\}$ by point (i) of Lemma 2.6, and we are allowed to identify $x - y$ with $x + \tilde{y}$ and $-y + x$ with $\tilde{y} + x$, which will turn to be very useful in various places later on.

3. THE DAVENPORT TRANSFORM REVISED

As mentioned in the introduction, Davenport's proof [6, Statement A] of Theorem 1 is a transformation proof. Loosely speaking, the idea is to map a pair (X, Y) of non-empty subsets of a *commutative* group $\mathbb{A} = (A, +, -, 0_A)$ to a new pair (X, Y') , which is smaller than (X, Y) in an appropriate sense, and specifically such that

$$|Y'| < |Y|, \quad |X + Y'| + |Y| \leq |X + Y| + |Y'|.$$

One then refers to (X, Y') as a Davenport transform of (X, Y) ; cf., e.g., [26, §3]. The construction implies that $X + 2Y \not\subseteq X + Y$ and $0_A \in Y$, to the effect that $|Y| \geq 2$.

As broadly expected, many difficulties arise when attempting to adapt the same approach to semigroups, and that all the more if these are non-commutative. Even the possibility of embedding a semigroup into a monoid does not resolve anything at all, since the fundamental problem is that, contrary to the case of groups, cardinality is not preserved “under difference”. To wit, if $\mathbb{A} = (A, +)$ is a unital semigroup with identity

0_A , X is a subset of \mathbb{A} , and a is an element of \mathbb{A} , then the cardinalities of X , $X - a$ and $-a + X$ can be greatly different from each other, even supposing that \mathbb{A} is cancellative; cf. point (ii) of Lemma 2.6. Thus, unless \mathbb{A} is a group in disguise or, more generally, embeds as a submonoid into a group, one is not allowed to assume, for instance, that $0_A \in Y$ by picking an arbitrary element $y_0 \in Y$ and by replacing (X, Y) with the pair $(X + y_0, -y_0 + Y)$; cf. Remark 3.

In fact, the primary goal of this section is to show that, in spite of such issues, Davenport's original ideas can be successfully extended to the more general setting of (cancellative) semigroups, and used to give a proof of Theorem 5.

To start with, suppose that $\mathbb{A} = (A, +)$ is a unital semigroup with identity 0_A and let X, Y be subsets of \mathbb{A} and $mX + 2Y \not\subseteq X + Y$ for some $m \in \mathbb{N}^+$. For brevity, define

$$Z := (mX + 2Y) \setminus (X + Y).$$

Our assumptions give $Z \neq \emptyset$. Thus, fix $z \in Z$, and take $x_z \in (m-1)X$ and $y_z \in Y$ such that $z \in x_z + X + Y + y_z$, where $0X := \{0_A\}$. Finally, set

$$(3) \quad \tilde{Y}_z := \{y \in Y : z \in x_z + X + Y + y\}, \quad Y_z := Y \setminus \tilde{Y}_z.$$

We refer to (X, Y_z) as a *generalized Davenport transform of (X, Y) (relative to z)* and, based on this notation, proceed on with the next proposition, intentionally organized in a list of properties argued starting from various “local” (rather than “global”) hypotheses, to remark differences with the “classical” Davenport transform and highlight which is used for which purpose.

Proposition 3.1. *If $Y_z \neq \emptyset$, then the triple (X, Y_z, \tilde{Y}_z) satisfies the following conditions:*

- (i) Y_z and \tilde{Y}_z are non-empty disjoint proper subsets of Y , and $\tilde{Y}_z = Y \setminus Y_z$.
- (ii) If \mathbb{A} is right cancellative, then $(x_z + X + Y_z) \cup (z - \tilde{Y}_z) \subseteq x_z + X + Y$.
- (iii) If $\langle Y \rangle_{\mathbb{A}}$ is commutative, then $(x_z + X + Y_z) \cap (z - \tilde{Y}_z) = \emptyset$.
- (iv) If \mathbb{A} is left cancellative, then $|z - \tilde{Y}_z| \geq |\tilde{Y}_z|$.
- (v) If \mathbb{A} is cancellative and $\langle Y \rangle_{\mathbb{A}}$ commutative, then $|X + Y| + |Y_z| \geq |X + Y_z| + |Y|$.

Proof. (i) By construction, $\tilde{y}_z \in Y_z$, that is Y_z and \tilde{Y}_z are both non-empty. Also, (3) gives that $Y_z, \tilde{Y}_z \subseteq Y$ and $Y_z \cap \tilde{Y}_z = \emptyset$, to the effect that $Y \setminus Y_z = Y \setminus (Y \setminus \tilde{Y}_z) = \tilde{Y}_z$ and $Y_z, \tilde{Y}_z \subsetneq Y$.

(ii) Since $Y_z \subseteq Y$ by point (i) above, $x_z + X + Y_z \subseteq x_z + X + Y$ by Lemma 2.2. On the other hand, if $w \in z - \tilde{Y}_z$ then there exists $y \in \tilde{Y}_z$ such that $z = w + y$. But $y \in \tilde{Y}_z$ implies by (3) that $z = \tilde{w} + y$ for some $\tilde{w} \in x_z + X + Y$, whence $w = \tilde{w}$ by right cancellativity, i.e. $w \in x_z + X + Y$.

(iii) Assume the contrary and let $w \in (x_z + X + Y_z) \cap (z - \tilde{Y}_z)$. There then exist $x \in X$, $y_1 \in Y_z$ and $y_2 \in \tilde{Y}_z$ such that $w = x_z + x + y_1$ and $z = w + y_2$. Using that $\langle Y \rangle_{\mathbb{A}}$ is commutative, this gives that $z = x_z + x + y_1 + y_2 = x_z + x + y_2 + y_1$, which implies that $y_1 \in \tilde{Y}_z$ by (3) since $Y_z, \tilde{Y}_z \subseteq Y$ by point (i). This is however absurd as $Y_z \cap \tilde{Y}_z = \emptyset$, again by point (i).

(iv) We have from (3) that for each $y \in \tilde{Y}_z$ there exists $w \in x_z + X + Y$ such that $z = w + y$, which yields that $w \in z - \tilde{Y}_z$. On the other hand, since \mathbb{A} is left cancellative,

it cannot happen that $w + y_1 = w + y_2$ for some $w \in \mathbb{A}$ and distinct $y_1, y_2 \in \tilde{Y}_z$. Thus, \tilde{Y}_z embeds as a set into $z - \tilde{Y}_z$, with the result that $|z - \tilde{Y}_z| \geq |\tilde{Y}_z|$.

(v) Since \mathbb{A} is cancellative and $X \neq \emptyset$ (otherwise $Z = \emptyset$), one has $|X + Y| \geq \max(|X|, |Y|)$ by point (ii) of Lemma 2.1 and point (ii) of Lemma 2.2. This implies the claim if Y is infinite, since then either $|X + Y| > |Y|$, and hence $|X + Y| + |Y_z| = |X| = |X + Y_z| + |Y|$, or $|X + Y| = |Y|$, and accordingly $|X + Y_z| + |Y_z| = |Y| = |X + Y_z| + |Y|$ (note that we are using here the AC). Thus, we are left with the case where Y is finite, for which the inclusion-exclusion principle, points (ii)-(iv) above, the same point (ii) of Lemma 2.1 and point (i) of Lemma 2.2 entail that

$$|X + Y| = |x_z + X + Y| \geq |x_z + X + Y_z| + |z - \tilde{Y}_z| = |X + Y_z| + |z - \tilde{Y}_z| \geq |X + Y_z| + |\tilde{Y}_z|.$$

But $\tilde{Y}_z = Y \setminus Y_z$ and $Y_z \subseteq Y$ by point (i), so that $|X + Y| \geq |X + Y_z| + |Y| - |Y_z|$. ■

Remark 5. To apply the generalized Davenport transform to the proof of Theorem 5, it will be enough to consider the case where $m = 1$, for which it is easily seen that $0_A \in Y_z$ if $0_A \in Y$ (we continue using the notation from above), as in the contrary case we would have $z \in X + Y$, contradicting the fact that $z \in (X + 2Y) \setminus (X + Y)$. However, it seems intriguing that the same machinery can be used, at least in principle, even if $m \geq 2$ in so far as one has a way to prove that $Y_z \neq \emptyset$, which is the reason why we decided to approach the subject as we have done.

4. THE PROOF OF THE MAIN THEOREM

Lemma 3.1 accounts for elementary properties of generalized Davenport transforms. It will serve here to establish the main contribution of the paper. For all practical intents, we remark that some results from Section 2, as basic as they are, will be used in the proof without explicit mention. This is especially the case of point (iii) of Lemma 2.1 and point (i) of Lemma 2.2.

Proof of Theorem 5. Since every semigroup embeds as a subsemigroup into any of its unitizations (unique up to a monoid isomorphism), and any unitization of a cancellative semigroup is cancellative in its own right, there is no loss of generality in assuming, as we do in the sequel in the light of Lemma 2.5 and Definition 1, that \mathbb{A} is unital.

Thus, suppose by contradiction that the theorem is false. Then, there must exist at least one pair (X, Y) of subsets of \mathbb{A} for which $|X + Y| < \Omega(X, Y)$, whence

$$(4) \quad 2 \leq |X|, |Y| < \infty.$$

In fact, if one of X or Y is empty then $|X + Y| = 0$, while if both X and Y are non-empty but one of them is a singleton or infinite then $|X + Y| = \max(|X|, |Y|)$. In both cases, Definition 1 gives that $|X + Y| = \Omega(X, Y)$, contradicting our assumptions. It follows from (2) and (4) that

$$(5) \quad |X + Y| < \sup_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0), \quad |X + Y| \leq |X| + |Y| - 2.$$

Again without loss of generality, we also assume that $|X| + |Y|$ is minimal among the pairs of subsets of \mathbb{A} for which (4) and (5) are presumed to hold. Now, since $|X + Y|$

is finite by (4) and point (i) of Lemma 2.3, one gets by (5) that Y^\times is non-empty and there exists $\tilde{y}_0 \in Y^\times$ such that

$$(6) \quad |X + Y| < \min_{y \in Y \setminus \{\tilde{y}_0\}} \text{ord}(y - \tilde{y}_0).$$

So letting 0_A denote the identity of \mathbb{A} and taking $W_0 := Y - \tilde{y}_0$ imply by (5) and (6) that

$$(7) \quad |X + W_0| < \min_{w \in W_0 \setminus \{0_A\}} \text{ord}(w), \quad |X + W_0| \leq |X| + |W_0| - 2,$$

as on the one hand $|Y - \tilde{y}_0| = |Y|$ and $|X + Y - \tilde{y}_0| = |X + Y|$ by point (ii) of Lemma 2.6, and on the other hand, $y \in Y \setminus \{\tilde{y}_0\}$ only if $y - \tilde{y}_0 \in (Y - \tilde{y}_0) \setminus \{0_A\}$ and $w \in (Y - \tilde{y}_0) \setminus \{0_A\}$ only if $w + \tilde{y}_0 \in Y \setminus \{\tilde{y}_0\}$ (see also Remark 4). We claim that

$$(8) \quad Z := (X + 2W_0) \setminus (X + W_0) \neq \emptyset.$$

For suppose the contrary. Then, $X + W_0 = X + \langle W_0 \rangle_{\mathbb{A}}$ by Lemma 2.4, with the result that

$$|X + W_0| = |X + \langle W_0 \rangle_{\mathbb{A}}| \geq |\langle W_0 \rangle_{\mathbb{A}}| \geq \max_{w \in W_0} \text{ord}(w) \geq \min_{w \in W_0 \setminus \{0_A\}} \text{ord}(w),$$

where we use, in particular, point (ii) of Lemma 2.2 for the first inequality and the fact that $|W_0| \geq 2$ for the last one. This is however in contradiction to (7), and hence (8) is proved.

Pick $z \in Z$ and let (X, W'_0) be a generalized Davenport transform of (X, W_0) relative to z . As Y generates a commutative subsemigroup of \mathbb{A} (by hypothesis), the same holds true with W_0 , by point (iii) of Lemma 2.6. Moreover, $0_A \in W_0$ (essentially by construction), and thus

$$(9) \quad 0_A \in W'_0 \neq \emptyset, \quad W'_0 \subsetneq W_0,$$

when taking into account Remark 5 and point (i) of Proposition 3.1. As a consequence, point (v) of the same Proposition 3.1 yields, together with (7), that

$$|X + W'_0| + |W_0| \leq |X + W_0| + |W'_0| \leq |X| + |W_0| - 2 + |W'_0|,$$

which ultimately means, since $|W_0| = |Y - \tilde{y}_0| = |Y| < \infty$ by (4) and the above, that

$$(10) \quad |X + W'_0| \leq |X| + |W'_0| - 2.$$

It follows from (9) that $1 \leq |W'_0| < |W_0|$, and indeed $|W'_0| \geq 2$, as otherwise (10) would give that $|X| = |X + W'_0| \leq |X| - 1$, which is absurd since $|X| < \infty$ by (4). To summarize,

$$(11) \quad 2 \leq |W'_0| < |W_0| < \infty.$$

Furthermore, (7) and (9) entail that

$$(12) \quad |X + W'_0| \leq |X + W_0| < \min_{w \in W'_0 \setminus \{0_A\}} \text{ord}(w),$$

where we use the elementary property that $\min(C_1) \leq \min(C_2)$ if C_1 and C_2 are sets of cardinal numbers with $C_2 \subseteq C_1$. Now, since $0_A \in W'_0{}^\times$, then (12) implies that

$$(13) \quad |X + W'_0| < \sup_{w_0 \in W'_0{}^\times} \min_{w \in W'_0 \setminus \{w_0\}} \text{ord}(w),$$

which gives, together with (4), (10) and (11), that $|X + W'_0| < \Omega(X, W'_0)$, contradicting the minimality of $|X| + |Y|$ in that $|W'_0| < |W_0| = |Y|$, and hence $|X| + |W'_0| < |X| + |Y|$. ■

5. SOME CONSEQUENCES

Now we prove a series of (rather immediate) corollaries, the first of which will confirm that Theorem 5 is actually both a generalization of Theorem 1 and a strengthening of Theorem 2, as already mentioned by the end of the introduction. We start with two simple lemmas.

Lemma 5.1. *Let $\mathbb{A} = (A, +)$ be a cancellative unital semigroup and Z a non-empty subset of \mathbb{A} such that $Z^\times \neq \emptyset$. Then $\sup_{z_0 \in Z^\times} \min_{z \in Z \setminus \{z_0\}} \text{ord}(z - z_0) \geq p(\mathbb{A})$.*

Proof. Pick $z_0 \in Z^\times$ using that $Z^\times \neq \emptyset$. If Z is a singleton, the assertion is trivial since then $\min_{z \in Z \setminus \{z_0\}} \text{ord}(z - z_0) = \infty$. Otherwise, let $z \in Z \setminus \{z_0\}$. Then, by the very definition of $p(\mathbb{A})$, one has $\text{ord}(z - z_0) \geq p(\mathbb{A})$, which is clearly enough to complete the proof. ■

Remark 6. In the case of groups, Lemma 5.1 obviously applies to *any* non-empty subset.

Lemma 5.2. *Let $\mathbb{A} = (A, +)$ be a cancellative semigroup and X, Y non-empty subsets of \mathbb{A} such that $Y^\times \neq \emptyset$. Then $\Omega(X, Y) \geq \min(p(\mathbb{A}), |X| + |Y| - 1)$.*

Proof. If (at least) one of X or Y is infinite, the assertion is trivial, in that $\Omega(X, Y) = |X| + |Y| - 1 = \max(|X|, |Y|)$. Otherwise, it follows from Lemma (5.1). ■

Corollary 5.1. *Let $\mathbb{A} = (A, +)$ be a cancellative semigroup and X, Y non-empty subsets of \mathbb{A} such that $\langle Y \rangle_{\mathbb{A}}$ is commutative and $Y^\times \neq \emptyset$. Then $|X + Y| \geq \min(p(\mathbb{A}), |X| + |Y| - 1)$.*

Proof. It is immediate by Lemma 5.2 and Theorem 5. ■

Remark 7. Corollary 5.1 contains Theorem 2 (and hence Theorem 1) as a special case.

The bound in Theorem 5 can be slightly strengthened in the case where both summands generate commutative subsemigroups.

Corollary 5.2. *Let $\mathbb{A} = (A, +)$ be a cancellative semigroup and X, Y subsets of \mathbb{A} such that $\langle X \rangle_{\mathbb{A}}$ is commutative. Then $|X + Y| \geq \Omega(Y, X)$.*

Proof. It is straightforward from point (i) of Lemma 2.1 and Theorem 5. ■

Corollary 5.3. *Let $\mathbb{A} = (A, +)$ be a cancellative semigroup and X, Y subsets of \mathbb{A} such that $\langle X \rangle_{\mathbb{A}}$ and $\langle Y \rangle_{\mathbb{A}}$ are both commutative. Then $|X + Y| \geq \max(\Omega(X, Y), \Omega(Y, X))$.*

Proof. It is a trivial consequence of Corollary 5.2 and Theorem 5. ■

Theorem 5 also allows for an alternative proof of I. Chowla's theorem for composite moduli [4]. Here, as in the statement of Theorem 6, we continue identifying a residue class with any of its representatives wherever it may be necessary.

Corollary 5.4. *Let $\mathbb{A} = (A, +)$ be a cancellative unital semigroup with identity 0_A and X, Y non-empty subsets of \mathbb{A} such that $\langle Y \rangle_{\mathbb{A}}$ is commutative and $0_A \in Y$. Then*

$$|X + Y| \geq \min \left(\min_{y \in Y \setminus \{0_A\}} \text{ord}(y), |X| + |Y| - 1 \right).$$

Proof. If one of X or Y is infinite, the claim is trivial, since then

$$|X + Y| = \max(|X|, |Y|) = |X| + |Y| - 1.$$

In all other cases, Theorem 5 implies that $|X + Y| \geq \min(\omega(Y), |X| + |Y| - 1)$, where

$$\omega(Y) := \sup_{y_0 \in Y^\times} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y) \geq \min_{y \in Y \setminus \{0_A\}} \text{ord}(y),$$

because $0_A \in Y^\times$. Clearly, this suffices to complete the proof. ■

Corollary 5.5. *Let $\mathbb{A} = (A, +, -, 0)$ be a group and X, Y non-empty subsets of \mathbb{A} such that $\langle Y \rangle_{\mathbb{A}}$ is commutative. Then, $|X + Y| \geq \min(\omega(Y), |X| + |Y| - 1)$, where*

$$\omega(Y) = \sup_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0),$$

and indeed $\omega(Y) = \max_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0)$ if \mathbb{A} is finite.

Proof. It is immediate from Theorem 5 since on the one hand \mathbb{A} being a group implies $Y = Y^\times$, and on the other a supremum over a finite set is a maximum (cf. Remark 1). ■

Theorem 7 (Chowla's theorem for composite moduli [4]). *Let $m \in \mathbb{N}^+$ and denote by \mathbb{A} the additive group $(\mathbb{Z}/m\mathbb{Z}, +, -, 0_m)$ of the integers modulo m . If X, Y are subsets of \mathbb{A} with $0_m \in Y$ and $\gcd(m, y) = 1$ for every $y \in Y \setminus \{0_m\}$, then $|X + Y| \geq \min(m, |X| + |Y| - 1)$.*

Proof. It is straightforward from Corollary 5.5 since $\text{ord}(y) = m$ for each non-zero residue class $y \in Y$, using that $\gcd(m, y) = 1$ by hypothesis. ■

And here is the proof of our generalization of Chowla's result:

Proof of Theorem 6. As a matter of fact, \mathbb{A} is a commutative finite group, for which one has especially that $\text{ord}(z - z_0) = m / \gcd(m, z - z_0)$ for all $z, z_0 \in \mathbb{A}$. Thus, Corollaries 5.3 and 5.5 entail that $|X + Y| \geq \min(\omega(Y), |X| + |Y| - 1)$, where

$$\omega(Y) = \max_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0) = m \cdot \max_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \frac{1}{\gcd(m, y - y_0)} = \delta_y^{-1} m.$$

Now in an entirely similar way, it is found, in the light of Corollary 5.2, that

$$|X + Y| \geq \min(\delta_x^{-1} m, |X| + |Y| - 1),$$

which is clearly enough to complete the proof by considering that $\delta_y = 1$ (respectively, $\delta_x = 1$) if there exists $y_0 \in Y$ (respectively, $x_0 \in X$) such that m is coprime with $y - y_0$ (respectively, with $x - x_0$) for every $y \in Y \setminus \{y_0\}$ (respectively, for every $x \in X \setminus \{x_0\}$). ■

Obviously, Theorem 6 contains Theorem 7 as a special case.

6. DISTINCT REPRESENTATIVES

As noticed by Ø. J. Rødseth in [26, §6], Davenport's paper [6] begins on the very same page of the 1935 volume of *Journal of the London Mathematical Society* which marks the end of the original article by P. Hall containing his well-known theorem about distinct representatives [13], herein referred to simply as Hall's theorem:

Hall's theorem. *For $k \in \mathbb{N}^+$ let S_1, S_2, \dots, S_k be sets. There then exist (pairwise) distinct elements s_1, s_2, \dots, s_k such that $s_i \in S_i$ and $s_i \neq s_j$ for $i \neq j$, if and only if the union of any h of these sets contains at least h elements for every $h = 1, 2, \dots, k$.*

Following [26, §6], we use Hall's theorem to say something on “how to spot”, by an algorithmic procedure, some of the elements of a sum-set. More precisely, suppose that $\mathbb{A} = (A, +)$ is a cancellative semigroup and let X, Y be non-empty finite subsets of \mathbb{A} such that

$$(14) \quad |X + Y| < \omega(Y) := \max_{y_0 \in Y^\times} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0).$$

Clearly, (14) implies that Y^\times is non-empty. Now, define $k := |X|$ and $\ell := |Y|$, and denote by x_1, x_2, \dots, x_k a numbering of X and by y_1, y_2, \dots, y_ℓ a numbering of Y . Then, consider the following matrix:

$$(15) \quad \alpha(X, Y) := \begin{bmatrix} x_1 + y_1 & x_1 + y_2 & \dots & x_1 + y_\ell \\ x_2 + y_1 & x_2 + y_2 & \dots & x_2 + y_\ell \\ \vdots & \vdots & \ddots & \vdots \\ x_k + y_1 & x_k + y_2 & \dots & x_k + y_\ell \end{bmatrix}.$$

Any element of $X + Y$ appears in some row or column of $\alpha(X, Y)$, and viceversa any entry of $\alpha(X, Y)$ is an element of $X + Y$. Also, Theorem 5, Remark 1 and our hypotheses give $|X + Y| \geq k + \ell - 1$. So it is natural to ask whether it is possible to argue anything about the actual positions where in the rectangular array (15) one can find $k + \ell - 1$ (distinct) elements of $X + Y$. Indeed, based on this notation, we can prove the following proposition, whose proof is almost identical to the one of [26, §6] (since it is short, we include it here for the sake of exposition):

Proposition 6.1. *Assume that $\langle Y \rangle_{\mathbb{A}}$ is commutative and let Z be any subset of $X + Y$ of size $\ell - 1$, for instance $Z = x_1 + \{y_1, y_2, \dots, y_{\ell-1}\}$. Then we can choose one element from each row of $\alpha(X, Y)$ so that Z and these elements form a subset of $X + Y$ of size $k + \ell - 1$.*

Proof. For each i let $Z_i := (x_i + Y) \setminus Z$ and note that Z_i is a set of elements in the i -th row of $\alpha(X, Y)$. Then, for all $h \in \mathbb{N}^+$ and pairwise distinct indices $i_1, i_2, \dots, i_h \in$

$\{1, 2, \dots, k\}$ one has $Z_{i_1} \cup Z_{i_2} \cup \dots \cup Z_{i_h} = (\{x_{i_1}, x_{i_2}, \dots, x_{i_h}\} + Y) \setminus Z$, with the result that

$$|Z_{i_1} \cup Z_{i_2} \cup \dots \cup Z_{i_h}| \geq |\{x_{i_1}, x_{i_2}, \dots, x_{i_h}\} + Y| - |Z| = h + \ell - 1 - (\ell - 1) = h,$$

thanks to Theorem 5 and the fact that $|\{x_{i_1}, x_{i_2}, \dots, x_{i_h}\} + Y| \leq |X + Y| \leq \omega(Y)$ by point (iii) of Lemma 2.1 and (14). It follows from Hall's theorem that we can find k distinct elements, one from each Z_i . Together with the $\ell - 1$ elements of Z , this provides a total amount of $k + \ell - 1$ elements of $X + Y$ since, by construction, $Z \cap Z_i = \emptyset$ for every i . \blacksquare

7. EXAMPLES

We conclude with a couple of representative examples showing that Theorem 5 can actually be greatly sharper than Theorem 2. Both of them are concerned with groups.

Example 1. Pick $k, q \in \mathbb{N}^+$ with q prime. Define $m := qk$, and let \mathbb{A} denote the additive group $(\mathbb{Z}/m\mathbb{Z}, +, -, 0_m)$ of the integers modulo m . Consider the subset

$$X := \{1 + ik : i = 0, 1, \dots, q - 1\}$$

of \mathbb{A} . Then, $2X = \{2 + ik : i = 0, 1, \dots, 2q - 2\}$ and $\Omega(X, X) = q$, while $p(\mathbb{A})$ is the smallest natural prime p dividing m . It follows that $p(\mathbb{A})$ is (much) smaller than $\Omega(X, X)$ if p is (much) smaller than q , while $|2X| = q + 1$ in all cases. Now taking the direct (group) product of many copies of \mathbb{A} , or more generally considering products relative to different moduli, yields other significant examples in the same spirit.

Example 2. Let A be a non-empty alphabet consisting of two or more letters. Pick $z_0 \in A$ and denote by \mathbb{A} the (multiplicatively-written) group with presentation $\langle A \mid z_0^2 = 1 \rangle_{\text{Grp}}$ (adding more relations in the presentation of \mathbb{A} yields a variety of similar examples). Let X, Y be non-empty finite subsets of \mathbb{A} with $Y \subseteq \langle z \rangle_{\mathbb{A}}$ for some $z \in A \setminus \{z_0\}$. Then, $\langle Y \rangle_{\mathbb{A}}$ is clearly commutative and

$$\sup_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \text{ord}(y - y_0) = |\mathbb{N}|,$$

to the effect that $|X + Y| \geq |X| + |Y| - 1$ by Corollary 5.5. On the other hand, one has that $\min(p(\mathbb{A}), |X| + |Y| - 1) = p(\mathbb{A}) = 2$ whenever $|X| + |Y| \geq 3$, with the result that the estimate on $|X + Y|$ provided by Theorem 2 is far too pessimistic, and definitely worse than the one furnished by Theorem 5, as soon as $|X| + |Y|$ is “large”.

8. ACKNOWLEDGMENTS

I am grateful to Andrea GAGNA (Università di Milano), Alain PLAGNE (École Polytechnique) and Carlo SANNA (Università di Torino) for helpful comments and fruitful discussions.

REFERENCES

- [1] Bibliography N. Alon, M. B. Nathanson and I. Ruzsa, ‘The polynomial method and restricted sums of congruence classes’, *J. Number Th.* 56 (1996) 404–417.
- [2] Bibliography A. L. Cauchy, ‘Recherches sur les nombres’, *J. École Polytech.* 9 (1813) 99–116; reproduced in *Oeuvres*, Série 2, Tome 1, 39–63.
- [3] Bibliography J. Cilleruelo, Y. O. Hamidoune and O. Serra, ‘Addition theorems in acyclic semigroups.’ In *Additive number theory* (Springer, 2010) 99–104.
- [4] Bibliography I. Chowla, ‘A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring’s problems’, *Proc. Indian Acad. Sc. (A)*, 2 (1935) 242–243.
- [5] Bibliography J. G. van der Corput, ‘On sets of integers I, II, III’, *Proc. Akad. Wet. Amsterdam* 50 (1947) 252–261, 340–350, 429–435.
- [6] Bibliography H. Davenport, ‘On the addition of residue classes’, *J. London Math. Soc.* 10 (1935) 30–32.
- [7] Bibliography H. Davenport, ‘A historical note’, *J. London Math. Soc.* 22 (1947) 100–101.
- [8] Bibliography F. J. Dyson, ‘A theorem on the densities of sets of integers’, *J. Lond. Math. Soc.* 20 (1945) 8–14.
- [9] Bibliography S. Eliahou and M. Kervaire, ‘Sumsets in vector spaces over finite fields’, *J. Number Th.* 71 (1998) 12–39.
- [10] Bibliography S. Eliahou and M. Kervaire, ‘Some extensions of the Cauchy-Davenport theorem’, *Electronic Notes in Discrete Mathematics* 28 (2007) 557–564.
- [11] Bibliography W. Feit and J. G. Thompson, ‘Solvability of groups of odd order’, *Pacific J. Math.* 13 (1963) 775–1029.
- [12] Bibliography A. Geroldinger, ‘Additive Group Theory and Non-unique Factorizations.’ In *Combinatorial Number Theory and Additive Group Theory* (Springer, 2009).
- [13] P. Hall, ‘On representatives of subsets’, *J. London Math. Soc.* 10 (1935), 26–30.
- [14] Bibliography Y. O. Hamidoune, ‘On a subgroup contained in words with a bounded length’, *Discrete Math.* 103 (1992) 171–176.
- [15] Bibliography P. Hegarty, ‘A Cauchy-Davenport Type Result for Arbitrary Regular Graphs’, *Integers* 11 No. 2 (Oct 2011) 227–235.
- [16] Bibliography J. M. Howie, *Fundamentals of semigroup theory* (Clarendon Press, 1995).
- [17] Bibliography G. Károlyi, ‘A compactness argument in the additive theory and the polynomial method’, *Discrete Math.* 302 Nos 1-3 (2005) 124–144.
- [18] Bibliography G. Károlyi, ‘The Cauchy-Davenport theorem in group extensions’, *L’Enseignement Mathématique* 51 (2005) 239–254.
- [19] Bibliography J. H. B. Kemperman, ‘On complexes in a semigroup’, *Indag. Math.* 18 (1956) 247–254.
- [20] Bibliography A. I. Mal’cev, ‘On the immersion of an algebraic ring into a field’, *Math. Annalen* 113 No. 1 (1937) 686–691.
- [21] Bibliography M. R. Murty and J. P. Whang, ‘The uncertainty principle and a generalization of a theorem of Tao’, *Linear Algebra and its Applications* 437 No. 1. (Jul 2012) 214–220.
- [22] Bibliography M. B. Nathanson, *Additive Number Theory. Inverse Problems and the Geometry of Sumsets* (GTM 165, Springer, 1996).
- [23] Bibliography J. E. Olson, ‘On the symmetric difference of two sets in a group’, *Europ. J. Comb.* 7 (1986) 43–54.
- [24] Bibliography S. S. Pillai, ‘On the addition of residue classes’, *Proc. Indian Acad. Sc. (A)*, 7 No. 1 (1938) 1–4.
- [25] Bibliography J. M. Pollard, ‘Addition properties of residue classes’, *J. London Math. Soc.* 11 No. 2 (1975) 147–152.
- [26] Bibliography Ø. J. Rødseth, ‘Sumsets mod p ’, *Skr. K. Nor. Vidensk. Selsk. (Trans. R. Norw. Soc. Sci. Lett.)* 4 (2006) 1–10.
- [27] Bibliography I. Z. Ruzsa, ‘Sumsets and structure.’ In *Combinatorial Number Theory and Additive Group Theory* (Springer, 2009).

- [28] BibliographyT. C. Tao, ‘An uncertainty principle for cyclic groups of prime order’, *Math. Res. Lett.* 12 No. 1 (2005) 121–127.
- [29] BibliographyS. Tringali, ‘Small doubling in ordered semigroups’, [arXiv:1208.3233](https://arxiv.org/abs/1208.3233) [math.CO, math.GT].
- [30] BibliographyA. G. Vosper, ‘The critical pairs of subsets of a group of prime order’, *J. London Math. Soc.* 31 (1956) 200–205.
- [31] BibliographyA. G. Vosper, ‘Addendum to “The critical pairs of subsets of a group of prime order”’, *J. London Math. Soc.* 31 (1956) 280–282.
- [32] BibliographyS. Yuzvinsky, ‘Orthogonal pairings of Euclidean spaces’, *Michigan Math. J.* 28 (1981) 109–119.

LABORATOIRE JACQUES-LOUIS LIONS, INSTITUTE DE MATHÉMATIQUE DE JUSSIEU, UNIVERSITÉ
PIERRE ET MARIE CURIE, 4 PLACE JUSSIEU, 75005 PARIS.

E-mail address: `tringali@ann.jussieu.fr`